

①⑨ RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
PARIS

①① N° de publication : 2 784 218
(à n'utiliser que pour les
commandes de reproduction)

②① N° d'enregistrement national : 98 12500

⑤① Int Cl⁷ : G 10 L 19/04 // G 10 L 101:10

①②

DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 06.10.98.

③① Priorité :

④③ Date de mise à la disposition du public de la
demande : 07.04.00 Bulletin 00/14.

⑤⑥ Liste des documents cités dans le rapport de
recherche préliminaire : *Se reporter à la fin du
présent fascicule*

⑥① Références à d'autres documents nationaux
apparentés :

⑦① Demandeur(s) : THOMSON CSF Société anonyme —
FR.

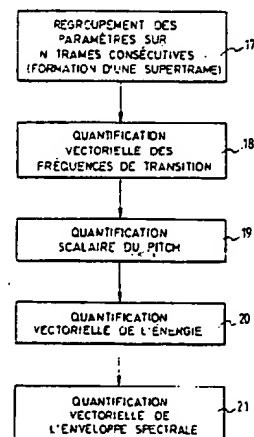
⑦② Inventeur(s) : GOURNAY PHILIPPE et CHARTIER
FREDERIC.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : THOMSON CSF.

⑤④ PROCÉDE DE CODAGE DE LA PAROLE A BAS DEBIT.

⑤⑦ Le procédé consiste à regrouper (17) les paramètres sur N trames consécutives pour former une super-trame, à effectuer une quantification vectorielle (18) des fréquences de transition du voisement au cours de chaque super-trame, en ne transmettant sans dégradation que les configurations les plus fréquentes et en remplaçant les configurations les moins fréquentes par la configuration la plus proche en terme d'erreur absolue parmi les plus fréquentes, à coder le pitch (19) en ne quantifiant scalairement qu'une seule valeur du pitch pour chaque super-trame, à coder l'énergie (20) en ne sélectionnant qu'un nombre réduit de valeurs en regroupant ces valeurs en sous paquets quantifiés par quantification vectorielle, à coder par quantification vectorielle (21) les paramètres d'enveloppe spectrale en ne sélectionnant qu'un nombre déterminé de filtres, les paramètres non transmis étant reconstruits par interpolation ou extrapolation à partir des paramètres des filtres transmis.
Applications: vocodeurs.



FR 2 784 218 - A1

Best Available Copy



La présente invention concerne un procédé de codage de la parole. Elle s'applique notamment à la réalisation de vocodeurs à très bas débit, de l'ordre de 1 200 bits par seconde et mis en oeuvre par exemple dans les communications par satellite, la téléphonie sur internet, les répondeurs statiques, les pageurs vocaux etc...

L'objectif de ces vocodeurs est de permettre de reconstruire un signal qui soit le plus proche possible au sens de la perception par l'oreille humaine du signal de parole d'origine, en utilisant un débit binaire le plus faible possible.

Pour atteindre cet objectif les vocodeurs utilisent un modèle totalement paramétré du signal de parole. Les paramètres utilisés concernent le voisement qui décrit le caractère périodique des sons voisés ou le caractère aléatoire de sons non voisés, la fréquence fondamentale des sons voisés encore connue sous le vocable anglo-saxon "PITCH", l'évolution temporelle de l'énergie ainsi que l'enveloppe spectrale du signal pour exciter et paramétrer les filtres de synthèse. Généralement le filtrage est réalisé par une technique de filtrage numérique à prédiction linéaire.

Ces différents paramètres sont estimés périodiquement sur le signal de parole, de une à plusieurs fois par trame de 10 à 30 ms, selon les paramètres et les codeurs. Ils sont élaborés au niveau d'un dispositif d'analyse et sont généralement transmis à distance en direction d'un dispositif de synthèse.

Le domaine du codage de la parole à bas débit a longtemps été dominé par un codeur à 2 400 bits/s connu sous la désignation LPC 10. Une description de ce codeur, ainsi que d'une variante à plus bas débit peut être trouvée dans les articles intitulés :

"Parameters and coding characteristics that must be common to assure interoperability of 2 400 bps linear predictive encoded speech", NATO Standard STANAG - 4198 - Ed 1, 13 February 1984 et dans l'article de MM. B. Mouy, D de la Nouë et G. Goudezeune, intitulé "NATO STANAG 4479 : A standard for a 800 bps vocoder and channel coding in HF-ECCM system", publié dans IEEE International Conference on

Acoustics, Speech, and Signal Processing, Detroit, May 1955, pp. 480-483.

Bien que parfaitement intelligible, la parole reproduite par ce vocodeur, est d'assez mauvaise qualité, de sorte que son usage est limité à des applications bien spécifiques, principalement professionnelles et militaires. Ces dernières années le domaine du codage de la parole à bas débit a connu un grand nombre d'innovations, grâce à l'introduction de nouveaux modèles connus respectivement sous les abréviations MBE, PWI et MELP.

Une description du modèle MBE peut être trouvée dans l'article de MM. D.W. Griffin and J.S. Lim, intitulé "Multiband Excitation Vocoders", publié dans la revue IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 36, n° 8, pp. 1223-1235, 1988.

Celle du modèle PWI peut être trouvée dans l'article de MM. W.B. Kleijn and J. Haogen, intitulé "Waveform Interpolation for Coding and Synthesis" dans la revue Speech Coding and Synthesis édité par W.B. Kleijn et K.K. Paliwal, Elsevier 1995.

Enfin, une description du modèle MELP peut être trouvée dans l'article de MM. L.M. Supplee, R.P. Cohn, J.S. Collura, and A.V. McCree, intitulé "MELP : The new federal standard at 2 400 bits/s, publié dans la revue IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, April 1997, pp. 1591 - 1594.

La qualité de la parole restituée par ces modèles à 2400 bits/s est devenue acceptable pour un grand nombre d'applications civiles et commerciales. Mais pour les débits inférieurs à 2 400 bits/s (typiquement 1 200 bits/s ou moins) la parole restituée présente une qualité insuffisante et pour pallier cet inconvénient d'autres techniques ont été mises en oeuvre. Une première technique est celle du vocodeur segmental, dont deux variantes sont celles décrites par MM. B. Mouy, P. de la Noue and G. Goudezeune déjà citée, et de celle décrite par M. Y. Shoham intitulée "Very low complexity interpolative speech coding at 1.2 to 2.4 K bps", publié dans IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, April 1997, pp 1599 - 1602.

Mais à ce jour, aucun vocodeur segmental n'a été jugé de qualité suffisante pour des applications civiles et commerciales.

Une deuxième technique est celle mise en oeuvre dans les vocodeurs phonétiques, qui combinent des principes de reconnaissance et de synthèse. L'activité dans ce domaine se situe plutôt au stade de la recherche fondamentale, les débits visés sont généralement très inférieurs à 1 200 bits/s (typiquement 50 à 200 bits/s) mais la qualité obtenue est plutôt mauvaise et il n'y a souvent pas de reconnaissance du locuteur. Une description de ces types de vocodeurs peut être trouvée dans l'article de MM. J. Cernocky, G. Baudoin, G. Chollet, ayant pour titre : "Segmental vocoder - Going beyond the phonetic approach" publié dans IEE International Conference on Acoustics, Speech, and Signal Processing, Seattle, May 12 - 15 1998, pp. 605 - 698.

Le but de l'invention est de pallier les inconvénients cités.

A cet effet, l'invention a pour objet un procédé de codage et de décodage de la parole pour les communications vocales utilisant un vocodeur à très bas débit comportant une partie analyse pour le codage et la transmission des paramètres du signal de parole et une partie synthèse pour la réception et le décodage des paramètres transmis et la reconstruction du signal de parole par utilisation de filtres de synthèse à prédiction linéaire du type consistant à analyser les paramètres, décrivant le pitch, la fréquence de transition de voisement, l'énergie, et l'enveloppe spectrale du signal de parole, en découpant le signal de parole en trames successives de longueur déterminée caractérisé en ce qu'il consiste à regrouper les paramètres sur N trames consécutives pour former une super-trame, à effectuer une quantification vectorielle des fréquences de transition du voisement au cours de chaque super-trame, en ne transmettant sans dégradation que les configurations les plus fréquentes et en remplaçant les configurations les moins fréquentes par la configuration la plus proche en terme d'erreur absolue parmi les plus fréquentes, à coder le pitch en ne quantifiant scalairement qu'une seule valeur pour chaque super-trame, à coder l'énergie en ne sélectionnant qu'un nombre réduit de valeurs en regroupant ces valeurs en sous paquets quantifiés par quantification vectorielle, les valeurs d'énergie non

transmises étant récupérées dans la partie synthèse par interpolation ou extrapolation à partir des valeurs transmises, à coder par quantification vectorielle les paramètres d'enveloppe spectrale pour l'encodage des filtres de synthèse à prédiction linéaire en ne sélectionnant qu'un nombre
5 déterminé de filtres, les paramètres non transmis étant reconstruits par interpolation ou extrapolation à partir des paramètres des filtres transmis.

D'autres caractéristiques et avantages de l'invention apparaîtront à l'aide de la description qui suit faite en regard des dossiers annexés qui représentent :

10 La figure 1 un modèle d'excitation mixte d'un vocodeur type HSX utilisé pour la mise en oeuvre de l'invention.

La figure 2 un schéma fonctionnel de la partie "analyse" d'un vocodeur de type HSX utilisé pour la mise en oeuvre de l'invention.

La figure 3 un schéma fonctionnel de la partie synthèse d'un
15 vocodeur de type HSX utilisé pour la mise en oeuvre de l'invention.

La figure 4 les étapes principales du procédé selon l'invention mises sous la forme d'un organigramme.

La figure 5 un tableau montrant la répartition des configurations des fréquences de transition de voisement pour trois
20 trames consécutives.

La figure 6 une table de quantification vectorielle des fréquences de transition de voisement utilisable pour la mise en oeuvre de l'invention.

La figure 7 une liste sous forme de tableau de schémas de
25 sélection et d'interpolation mise en oeuvre dans l'invention pour le codage de l'énergie du signal de parole.

La figure 8 une liste sous forme d'un tableau de schémas de sélection et d'interpolation/extrapolation pour l'encodage des filtres LPC à prédiction linéaire.

30 La figure 9 un tableau d'allocation des bits nécessaires au codage d'un vocodeur de type HSX à 1 200 bits/s selon l'invention.

Le procédé selon l'invention met en oeuvre un vocodeur de type connu sous l'abréviation anglo-saxonne HSX de "Harmonic

Stochastic Excitation", comme base pour la réalisation d'un vocodeur de bonne qualité à 1 200 bits/s.

Une description de ce type de vocodeur peut être trouvée dans l'article de MM. C. Laflamme, R. Salami, R. Matmti et J.P. Adoul, ayant
5 pour titre "Harmonic Stochastic Excitation (HSX) speech coding below 4 k.bits/s" et publié dans IEEE International Conference on Acoustics, and Signal Processing, Atlanta, May 1996, pp.204- 207.

Le procédé selon l'invention porte sur l'encodage des paramètres qui permet de reproduire au mieux avec un minimum de débit
10 toute la complexité du signal de parole.

Comme schématisé à la figure 1 un vocodeur HSX est un vocodeur à prédiction linéaire qui utilise dans sa partie synthèse un modèle d'excitation mixte simple, dans lequel un train d'impulsion périodique excite les fréquences basses et un niveau de bruit excite les
15 fréquences hautes d'un filtre LPC de synthèse. La figure 1 décrit le principe de génération de l'excitation mixte qui comporte deux voies de filtrage. La première voie 1₁ est excitée par un train d'impulsion périodique effectue un filtrage passe bas et la deuxième voie 1₂ excitée par un signal de bruit stochastique effectue un filtrage passe haut. La
20 fréquence de coupure ou de transition f_c des filtres des deux voies est la même et a une position variable dans le temps. Les filtres des deux voies sont complémentaires. Un sommateur 2 additionne les signaux fournis par les deux voies. Un amplificateur 3 de gain g ajuste le gain de la première voie de filtrage pour que le signal d'excitation obtenu en sortie
25 du sommateur 2 soit à spectre plat.

Un diagramme fonctionnel de la partie analyse du vocodeur est représenté à la figure 2. Pour effectuer cette analyse le signal de parole est d'abord filtré par un filtre passe haut 4 pour être ensuite segmenté en trames de 22,5 ms, comportant 180 échantillons prélevés à la fréquence
30 8 KHz. Deux analyses par prédiction linéaire sont effectuées en 5 sur chacune des trames. Aux étapes 6 et 7 le signal semi blanchi obtenu est filtré en quatre sous bandes. Un suiveur de pitch 8 robuste exploite la première sous bande. La fréquence de transition f_c entre la bande de fréquence basse des sons voisés et la bande de fréquence haute des sons

non voisés est déterminée par le taux de voisement mesuré en 9 dans les quatre sous bandes. Enfin, l'énergie est mesurée et codée à l'étape 10 de manière pitch-synchrone, 4 fois par trame.

Comme les performances du suiveur de pitch et de l'analyseur de voisement 9 peuvent être grandement améliorées lorsque leur décision est retardé d'une trame, les paramètres résultant, coefficients des filtres de synthèse, pitch, voisement, fréquence de transition et énergie sont codés avec une trame de retard.

Dans la partie synthèse du vocodeur HSX qui est représenté à la figure 3, le signal d'excitation du filtre de synthèse est formé de la manière déjà représentée à la figure 1 par la somme d'un signal harmonique et d'un signal aléatoire dont les enveloppes spectrales sont complémentaires. La composante harmonique est obtenue en passant un train d'impulsions à la période pitch dans un filtre passe bande précalculé 11. La composante aléatoire est obtenue à partir d'un générateur 12 combinant une transformée de Fourier inverse et un recouvrement temporel. Le filtre LPC de synthèse 14 est interpolé 4 fois par trame. Le filtre perceptuel 15 couplé en sortie de filtre 14 permet d'obtenir une meilleure restitution des caractéristiques nasales du signal de parole originel. Enfin le dispositif de contrôle automatique de gain permet d'assurer que l'énergie pitch-synchrone du signal de sortie est égale à celle qui a été transmise.

Avec un débit aussi bas que 1200 bits/s, il n'est pas possible d'encoder de façon précise toutes les 22,5 ms les 4 paramètres pitch, fréquence de transition de voisement, énergie et coefficients des deux filtres LPC à 10 coefficients par trame.

Pour exploiter au mieux les caractéristiques temporelles de l'évolution des paramètres qui comportent des périodes de stabilité entrecoupées de variations rapides, le procédé selon l'invention se déroule en cinq étapes principales référencées de 17 à 21 sur la figure 4. L'étape 17 regroupe les trames vocodeurs par N trames pour former une super trame. A titre indicatif une valeur de N égale à 3 peut être choisie car elle réalise un bon compromis entre la réduction possible du débit binaire et le retard introduit par le procédé de quantification. D'autre part,

elle est compatible avec les techniques d'entrelacement et de codage correcteur d'erreurs actuelles.

La fréquence de transition de voisement est codée à l'étape 18 par quantification vectorielle en utilisant uniquement quatre valeurs de fréquence, 0,750,2000 et 3625 HZ par exemple. Dans ces conditions 6 bits à raison de 2 bits par trame sont suffisants pour coder chacune des fréquences et transmettre exactement la configuration de voisement des trois trames d'une super trame. Toutefois comme certaines configurations de voisement ne se reproduisent que très rarement, on peut considérer qu'elles ne sont pas forcément caractéristiques de l'évolution du signal de parole normal, car elles ne semblent pas participer à l'intelligibilité, ni à la qualité de la parole restituée. C'est le cas par exemple lorsque une trame est totalement voisée de 0 Hz jusqu'à 3 625 Hz et qu'elle est comprise entre deux trames totalement non voisées.

Le tableau de la figure 5 retrace une répartition de configuration de voisement sur trois trames successives, calculées sur une base de données de 123 158 trames de parole. Dans ce tableau les 32 configurations les moins fréquentes comptent pour seulement 4% de toutes les trames, partiellement ou totalement voisées. La dégradation obtenue en remplaçant chacune de ces configurations par la plus proche, en terme d'erreur absolue, des 32 configurations les plus représentées est imperceptible. Ceci montre qu'il est possible d'économiser un bit en quantifiant vectoriellement la fréquence de transition de voisement sur une super trame. Une quantification vectorielle des configurations de voisement est montrée dans le tableau référencé 22 sur la figure 6. Le tableau 22 est organisé de sorte que l'erreur quadratique moyenne produite par une erreur sur un bit d'adressage soit minimale.

Le codage du pitch s'exécute à l'étape 19. Il met en oeuvre un quantificateur scalaire sur 6 bits, avec une plage d'échantillons de 16 à 148, et un pas de quantification uniforme sur une échelle logarithmique. Une seule valeur est transmise pour trois trames consécutives. Le calcul de la valeur à quantifier à partir des trois valeurs de pitch et la procédure permettant de récupérer les trois valeurs de pitch à partir de la valeur

quantifiée, diffèrent selon la valeur des fréquences de transition de voisement de l'analyse. Le processus est le suivant:

1. Lorsque aucune trame n'est voisée, les 6 bits sont positionnés à zéro, le pitch décodé est fixé à une valeur arbitraire soit, par exemple, à 45 échantillons pour chacune des trames de la super trame.

2. Lorsque la dernière trame de la super-trame précédente et les trois trames de la super trame courante sont voisées, c'est à dire, lorsque la fréquence de transition de voisement est supérieure strictement à zéro, la valeur quantifiée est la valeur du pitch de la dernière trame de la super trame courante qui est alors considérée comme une valeur cible. Au décodeur la valeur décodée du pitch pour la troisième trame de la super-trame courante est la valeur cible quantifiée, et les valeurs du pitch décodés pour les deux premières trames de la super-trame courante sont récupérées par interpolation linéaire entre la valeur transmise pour la super-trame précédente et la valeur cible quantifiée.

3. Pour toutes les autres configurations de voisement, c'est la valeur pondérée du pitch sur les trois trames de la super-trame courante qui est quantifiée. Le facteur de pondération est proportionnel à la fréquence de transition de voisement pour la trame considérée suivant la relation :

$$\text{Valeur Moyenne Pondérée} = \frac{\sum_{i=1-3} \text{Pitch}(i) * \text{voisement}(i)}{\sum_{i=1-3} \text{voisement}(i)}$$

Au décodeur la valeur du pitch décodée pour les trois trames de la super-trame courante est égale à la valeur moyenne pondérée quantifiée.

De plus dans les cas 2 et 3, un léger trémolo est appliqué systématiquement aux valeurs du pitch utilisées en synthèse pour les trames 1, 2 et 3 pour améliorer le naturel de la parole restituée en évitant la génération de signaux trop fortement périodiques, suivant par exemple les relations :

Pitch utilisé (1) = 0,995 * Pitch Décodé (1)
 Pitch utilisé (2) = 1,005 * Pitch Décodé (2)
 Pitch utilisé (3) = 1,000 * Pitch Décodé (3)

L'intérêt de procéder à une quantification scalaire des valeurs de pitch est qu'il limite le problème de propagation des erreurs sur le train binaire. De plus les schémas de codage 2 et 3 sont suffisamment proches l'un de l'autre pour être insensibles aux mauvais décodages de la
5 fréquence de voisement.

L'encodage de l'énergie est effectué à l'étape 20. Il a lieu de la façon représentée dans le tableau référencé 23 sur la figure 7 en utilisant une méthode de quantification vectorielle du type de celle décrite dans l'article de RM Gray, ayant pour titre "Vector Quantization", publié dans
10 la revue IEEE ASP Magazine, vol. 1, pp 4-29, avril 1984. Douze valeurs d'énergie numérotées de 0 à 11 sont calculées à chaque super-trame par la partie analyse et seulement six valeurs d'énergie parmi les douze sont transmises. Ceci conduit à construire deux vecteurs de trois valeurs par la partie analyse. Chaque vecteur est quantifié sur six bits. Deux bits sont
15 utilisés pour transmettre le numéro de schéma de sélection utilisé. Lors du décodage dans la partie synthèse, les valeurs de l'énergie qui n'ont pas été quantifiées sont récupérées par interpolation.

Seuls quatre schémas de sélection sont autorisés comme le montre le tableau de la figure 7. Ces schémas sont optimisés afin
20 d'encoder au mieux, soit les vecteurs de 12 énergies stables, soit ceux pour lesquels l'énergie varie rapidement au cours des trames 1, 2, et 3. Dans la partie analyse le vecteur d'énergie est encodé selon chacun des quatre schémas, et le schéma effectivement transmis est celui qui minimise l'erreur quadratique totale.

25 Dans ce processus les bits donnant le numéro du schéma transmis ne sont pas considérés comme sensibles, puisque qu'une erreur sur leur valeur ne fait qu'altérer légèrement l'évolution temporelle de la valeur de l'énergie. De plus la table de quantification vectorielle des énergies est organisée pour que l'erreur quadratique moyenne produite
30 par une erreur sur un bit d'adressage soit minimale.

Le codage des coefficients modélisant l'enveloppe du signal de parole a lieu par quantification vectorielle à l'étape 21. Ce codage permet de déterminer les coefficients des filtres numériques utilisés dans la partie synthèse. Six filtres LPC à 10 coefficients numérotés de 0 à 5 sont

calculés à chaque super-trame par la partie analyse et seulement 3 filtres parmi les 6 sont transmis. Les six vecteurs sont transformés en six vecteurs de 10 paires de raies spectrales LSF suivant par exemple le processus décrit dans l'article de M F. ITAKURA, intitulé "Line Spectrum
5 Représentation of Linear Predictive Coefficients" et publié dans le Journal Acoustique Society America, vol.57, P.35, 1975. Les paires de raies spectrales sont encodées par une technique similaire à celle mise en oeuvre pour le codage de l'énergie. Le processus consiste à sélectionner trois filtres LPC, et à quantifier chacun des vecteurs sur 18 bits en
10 utilisant par exemple un quantificateur vectoriel prédictif en boucle ouverte, avec un coefficient de prédiction égal à 0,6, de type SPLIT -VQ portant sur deux sous-paquets de 5 LSF consécutives auxquels il est alloué à chacun 9 bits. Deux bits sont utilisés pour transmettre le numéro du schéma de sélection utilisé. Au niveau du décodeur lorsqu'un filtre
15 LPC n'est pas quantifié, sa valeur est estimée à partir de celle des filtres LPC quantifiés par interpolation linéaire par exemple, ou par extrapolation par duplication par exemple du filtre LPC précédent. A titre d'exemple un processus de quantification vectorielle par paquets pourra être constitué de la façon décrite dans l'article de MM K.K. PALIWAL, BS. ATAL, ayant
20 pour titre "Efficient Vector Quantization of LPC Parameters at 24 bits/frame" et publié dans IEEE transaction on Speech and Audio Processing, Vol.1, Janvier 1993.

Comme indiqué dans le tableau référencé 24 sur la figure 8, seuls quatre schémas de sélection sont autorisés. Ces schémas
25 permettent de coder au mieux, soit les zones pour lesquelles l'enveloppe spectrale est stable, soit les zones pour lesquelles l'enveloppe spectrale varie rapidement au cours des trames 1, 2, ou 3. L'ensemble des filtres LPC est alors codé selon chacun des quatre schémas, et le schéma effectivement transmis est celui qui minimise l'erreur quadratique totale.

30 De manière similaire au codage de l'énergie, les bits donnant le numéro du schéma ne sont pas à considérer comme sensibles, puisque une erreur sur leur valeur ne fait qu'altérer légèrement l'évolution temporelle des filtres LPC. De plus les tables de quantification vectorielle des LSF sont organisées dans la partie synthèse de sorte que l'erreur

quadratique moyenne produite par une erreur sur un bit d'adressage soit minimum.

L'allocation des bits pour la transmission des paramètres LSF, de l'énergie, du pitch et du voisement qui résulte de la méthode de codage mise en oeuvre par l'invention est représentée dans le tableau de la figure 9 dans le cadre d'un vocodeur à 1200 bits/s dans lequel les paramètres sont codés toutes les 67,5 ms; 81 bits étant disponibles à chaque super trame pour encoder les paramètres du signal. Ces 81 bits se décomposent en 54 bits LSF, 2 bits pour la décimation du schéma des LSF, 2 fois 6 bits pour l'énergie, 6 bits pour le pitch et 5 bits pour le voisement.

REVENDICATIONS

1. Procédé de codage et de décodage de la parole pour les
5 communications vocales utilisant un vocodeur à très bas débit
comportant une partie analyse (4,...10) pour le codage et la transmission
des paramètres du signal de parole et une partie synthèse (11,...16) pour
la réception et le décodage des paramètres transmis et la reconstruction
10 du signal de parole par utilisation de filtres de synthèse à prédiction
linéaire du type consistant à analyser les paramètres, décrivant le pitch
(8), la fréquence de transition de voisement (9), l'énergie (10), et
l'enveloppe spectrale (5) du signal de parole, en découpant le signal de
parole en trames successives de longueur déterminée caractérisé en ce
15 qu'il consiste à regrouper (17) les paramètres sur N trames consécutives
pour former une super-trame, à effectuer une quantification vectorielle
(18) des fréquences de transition du voisement au cours de chaque
super-trame, en ne transmettant sans dégradation que les configurations
les plus fréquentes et en remplaçant les configurations les moins
20 fréquentes par la configuration la plus proche en terme d'erreur absolue
parmi les plus fréquentes, à coder le pitch (19) en ne quantifiant
scalairement qu'une seule valeur du pitch pour chaque super-trame, à
coder l'énergie (20) en ne sélectionnant qu'un nombre réduit de valeurs
en regroupant ces valeurs en sous paquets quantifiés par quantification
vectorielle, les valeurs d'énergie non transmises étant récupérées dans la
25 partie synthèse par interpolation ou extrapolation à partir des valeurs
transmises, à coder par quantification vectorielle (21) les paramètres
d'enveloppe spectrale pour l'encodage des filtres de synthèse à prédiction
linéaire en ne sélectionnant qu'un nombre déterminé de filtres, les
paramètres non transmis étant reconstruits par interpolation ou
30 extrapolation à partir des paramètres des filtres transmis.

2. Procédé selon la revendication 1 caractérisé en ce que la
valeur quantifiée du pitch est soit la dernière valeur du pitch des zones
stables entièrement voisées, soit une valeur moyenne pondérée par la

fréquence de transition de voisement dans les zones qui ne sont pas entièrement voisées.

3. Procédé selon la revendication 2 caractérisé en ce qu'il consiste lorsque la valeur de pitch est la dernière d'une super-trame, à
5 reconstituer les autres valeurs par interpolation.

4. Procédé selon la revendication 3 caractérisé en ce que la valeur du pitch utilisée dans la partie synthèse est celle du pitch décodé modifié par un coefficient de multiplication pour produire un léger trémolo dans la parole reconstituée.

10 5. Procédé selon l'une quelconque des revendications 1 à 4 caractérisé en ce que les paramètres sont regroupés sur un nombre $N=3$ de trames consécutives.

6. Procédé selon la revendication 5 caractérisé en ce que les fréquences de voisement sont au nombre de 4 et sont codées
15 vectoriellement à l'aide d'une table de quantification (22) comportant 32 configurations de fréquences groupées par 3.

7. Procédé selon l'une quelconque des revendications 5 et 6 caractérisé en ce qu'il consiste à mesurer l'énergie 4 fois par trame, seulement 6 valeurs parmi les 12 d'une super-trame étant transmises
20 (23) sous la forme de deux vecteurs de 3 valeurs.

8. Procédé selon la revendication 7 caractérisé en ce qu'il consiste à coder l'énergie (23) suivant quatre schémas regroupant chacun deux vecteurs, un premier schéma lorsque les douze vecteurs d'énergie dans la super-trame sont stables, les schémas restants étant définis pour
25 chacune des trames, et à transmettre le schéma qui minimise l'erreur quadratique totale.

9. Procédé selon la revendication 8 caractérisé en ce que :

- dans le premier schéma seules les valeurs d'énergie numérotées 1, 3, et 5 du premier vecteur et celles numérotées 7, 9, 11
30 du deuxième vecteur sont transmises,
- dans le deuxième schéma seules les valeurs d'énergies numérotées 0, 1, et 2 du premier vecteur et celles numérotées 3, 7, et 11 du deuxième vecteur sont transmises,

- dans le troisième schéma seules les valeurs d'énergies numérotées 1, 4 5 du premier vecteur et celles numérotées 6, 7, et 11 du deuxième vecteur sont transmises,

- et dans le quatrième schéma seules les valeurs d'énergies numérotées 2, 5 et 8 du premier vecteur et celles numérotées 9, 10 et 11 du deuxième vecteur sont transmises.

10. Procédé selon l'une quelconque des revendications 1 à 9 caractérisé en ce qu'il consiste à effectuer la sélection des paramètres d'encodage des filtres de prédiction linéaire suivant quatre schémas pour encoder au mieux soit les zones pour lesquelles l'enveloppe spectrale est stable, soit les zones pour lesquelles l'enveloppe spectrale varie rapidement au cours des trames 1, 2, ou 3 d'une super trame.

11. Procédé selon la revendication 10 caractérisé en ce qu'il consiste à utiliser (24) dans la partie synthèse 6 filtres à prédiction linéaire à 10 coefficients numérotés de 0 à 5 et à transmettre :

- dans un premier schéma que les coefficients des filtres 1, 3, et 5 lorsque l'enveloppe spectrale est stable,

- dans un deuxième schéma correspondant à la première trame que les coefficients des filtres 0, 1 et 4,

20 - dans un troisième schéma correspondant à la deuxième trame que les coefficients des filtres 2, 3 et 5,

- dans un quatrième schéma correspondant à la troisième trame que les coefficients des filtres 1, 4 et 5,

le schéma effectivement transmis étant celui qui minimise l'erreur quadratique totale, les coefficients des filtres non transmis étant calculés dans la partie synthèse par interpolation ou extrapolation.

12. Procédé selon l'une quelconque des revendications 1 à 11 caractérisé en ce que les coefficients LSF des filtres de synthèse sont codés sur un nombre de 54 bits auquel est ajouté deux bits pour la transmission des schémas de décimation, l'énergie est codée avec un nombre de 2 fois 6 bits auquel est ajouté 2 bits pour la transmission des schémas de décimation, le pitch est codé sur un nombre de 6 bits et la fréquence de transition de voisement est codée sur un nombre de 5 bits soit au total 81 bits pour des super-trames de 67,5 ms.

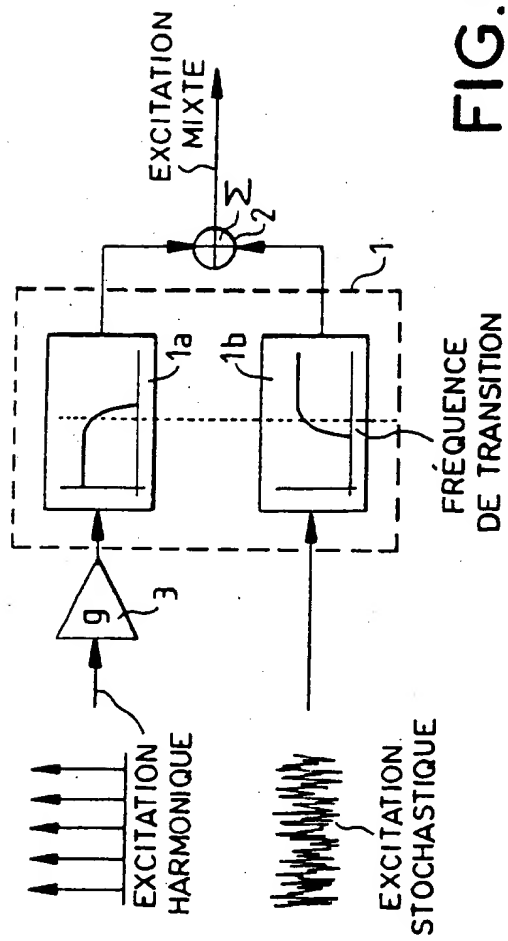


FIG. 1

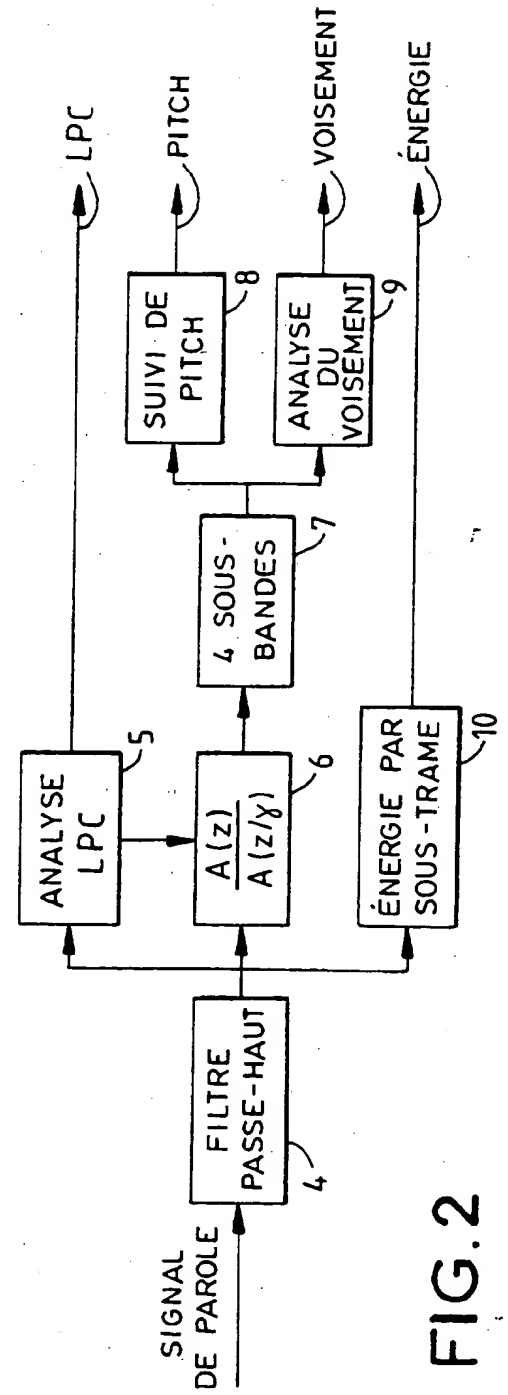


FIG. 2

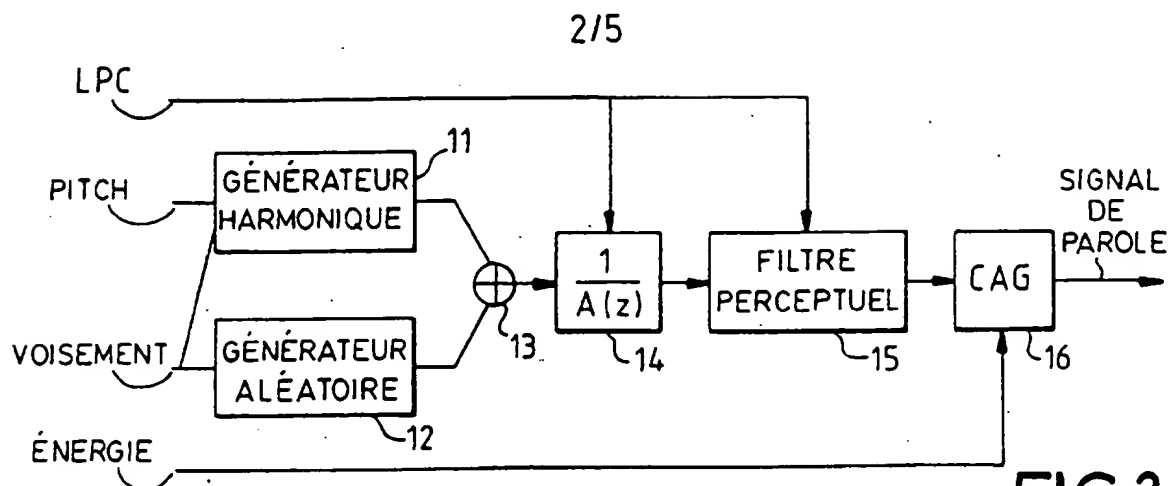


FIG.3

| CONFIGURATION DE VOISEMENT | NB. DE SUPERTRAMES |
|---|--------------------|
| TOTALEMENT NON-VOISÉ (Y COMPRIS LES SILENCES) | 55.585 |
| TOTALEMENT VOISÉ | 34.586 |
| PARTIELLEMENT NON-VOISÉ (30 PLUS FRÉQUENTES) | 30.273 |
| PARTIELLEMENT NON-VOISÉ (32 MOINS FRÉQUENTES) | 2.714 |
| TOTAL SUPERTRAMES : | 123.158 |

FIG.5

| NOM DU SCHÉMA | VECTEUR 1 | VECTEUR 2 | VALEURS INTERPOLÉES |
|---------------|-----------|-----------|---------------------|
| STABLE | 1, 3, 5 | 7, 9, 11 | 0, 2, 4, 6, 8, 10 |
| TRAME 1 | 0, 1, 2 | 3, 7, 11 | 4, 5, 6, 8, 9, 10 |
| TRAME 2 | 1, 4, 5 | 6, 7, 11 | 0, 2, 3, 8, 9, 10 |
| TRAME 3 | 2, 5, 8 | 9, 10, 11 | 0, 1, 3, 4, 6, 7 |

23

FIG.7

| NOM DU SCHÉMA | LPC QUANTIFIÉ | LPC INTERPOLÉ | LPC EXTRAPOLÉ |
|---------------|---------------|---------------|---------------|
| STABLE | 1, 3, 5 | 0, 2, 4 | — |
| TRAME 1 | 0, 1, 4 | 2, 3 | 5 |
| TRAME 2 | 2, 3, 5 | 0, 1, 4 | — |
| TRAME 3 | 1, 4, 5 | 0, 2, 3 | — |

24

FIG.8

3/5

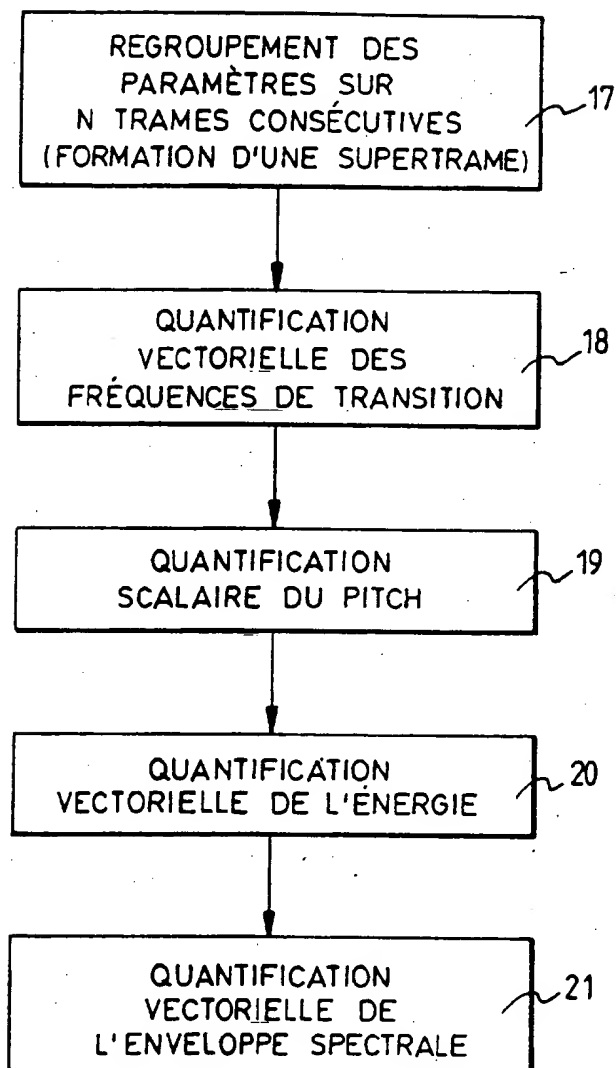


FIG. 4

4/5

22 ↗

| | | |
|------|------|------|
| 0 | 750 | 3625 |
| 0 | 3625 | 3625 |
| 0 | 2000 | 3625 |
| 750 | 3625 | 3625 |
| 0 | 0 | 3625 |
| 750 | 2000 | 3625 |
| 750 | 750 | 3625 |
| 2000 | 2000 | 3625 |
| 0 | 0 | 2000 |
| 750 | 750 | 750 |
| 750 | 750 | 0 |
| 2000 | 2000 | 750 |
| 0 | 0 | 750 |
| 0 | 0 | 0 |
| 750 | 0 | 0 |
| 2000 | 750 | 0 |
| 2000 | 2000 | 2000 |
| 3625 | 3625 | 2000 |
| 2000 | 3625 | 3625 |
| 3625 | 3625 | 3625 |
| 2000 | 750 | 750 |
| 3625 | 2000 | 750 |
| 3625 | 2000 | 2000 |
| 3625 | 2000 | 3625 |
| 2000 | 2000 | 0 |
| 3625 | 2000 | 0 |
| 3625 | 3625 | 0 |
| 3625 | 3625 | 750 |
| 2000 | 0 | 0 |
| 3625 | 0 | 0 |
| 3625 | 750 | 0 |
| 3625 | 750 | 750 |

FIG. 6

| PARAMÈTRE | NB. BITS |
|--------------------------------|----------|
| LSFs | 54 |
| SCHÉMA DE DÉCIMATION (LSFs) | 2 |
| ÉNERGIE | 2*6 |
| SCHÉMA DE DÉCIMATION (ÉNERGIE) | 2 |
| PITCH | 6 |
| VOISEMENT | 5 |
| TOTAL BITS / 67.5ms | 81 |

FIG.9

INSTITUT NATIONAL
de la
PROPRIETE INDUSTRIELLE

**RAPPORT DE RECHERCHE
PRELIMINAIRE**
établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 565674
FR 9812500

| DOCUMENTS CONSIDERES COMME PERTINENTS | | Revendications concernées de la demande examinée |
|---|---|--|
| Catégorie | Citation du document avec indication, en cas de besoin, des parties pertinentes | |
| A | <p>ATKINSON I ET AL: "High quality split band LPC vocoder operating at low bit rates"</p> <p>1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (CAT. NO.97CB36052), 1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, MUNICH, GERMANY, 21-24 APRIL 1997, pages 1559-1562 vol.2, XP002105943</p> <p>ISBN 0-8186-7919-0, 1997, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, USA</p> <p>* alinéa 5 *</p> | 1 |
| D,A | <p>MOUY B ET AL: "NATO STANAG 4479: A STANDARD FOR AN 800 BPS VOCODER AND CHANNEL CODING IN HF-ECCM SYSTEM" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), DETROIT, MAY 9 - 12, 1995 SPEECH, vol. 1, 9 mai 1995, pages 480-483, XP000658035</p> <p>INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS</p> <p>* alinéa II.2 *</p> <p>* alinéa II.3 *</p> | 1 |
| | | <p>DOMAINES TECHNIQUES RECHERCHES (Int.CL.6)</p> <p>G10L</p> |
| Date d'achèvement de la recherche | | Examineur |
| 16 juin 1999 | | Krembel, L |
| <p>CATEGORIE DES DOCUMENTS CITES</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant</p> | | |

2

EPO FORM 1503 02.82 (P04C13)

INSTITUT NATIONAL
de la
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE
PRELIMINAIRE

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 565674
FR 9812500

| DOCUMENTS CONSIDERES COMME PERTINENTS | | Revendications concernées de la demande examinée |
|--|---|---|
| Catégorie | Citation du document avec indication, en cas de besoin, des parties pertinentes | |
| A | YELDENER S ET AL: "A mixed sinusoidally excited linear prediction coder at 4 kb/s and below" PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, ICASSP '98 (CAT. NO.98CH36181), PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, SEATTLE, WA, USA, 12-1, pages 589-592 vol.2, XP002105944 ISBN 0-7803-4428-6, 1998, New York, NY, USA, IEEE, USA * page 591, colonne 1, ligne 19 - ligne 41 * * figure 4 * | 1 |
| A | US 5 774 837 A (AGUILAR JOSEPH GERARD ET AL) 30 juin 1998 * colonne 16, ligne 15 - ligne 47 * | 1 |
| A | MCCREE A V ET AL: "A MIXED EXCITATION LPC VOCODER WITH FREQUENCY-DEPENDENT VOICING STRENGTH" SPEECH AND AUDIO CODING FOR WIRELESS AND NETWORK APPLICATIONS, pages 259-264, XP000470449 ATAL B S CUPERMAN V; GERSHO A * figure 1 * | 1 |
| | | DOMAINES TECHNIQUES RECHERCHES (Int.CL.6) |
| | | |
| Date d'achèvement de la recherche | | Examineur |
| 16 juin 1999 | | Krembel, L |
| CATEGORIE DES DOCUMENTS CITES | | |
| <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant</p> | | |

2

EPO FORM 1503 03 82 (P04C13)

INSTITUT NATIONAL
de la
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE
PRELIMINAIRE
établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 565674
FR 9812500

| DOCUMENTS CONSIDERES COMME PERTINENTS | | Revendications concernées de la demande examinée |
|---|---|---|
| Catégorie | Citation du document avec indication, en cas de besoin, des parties pertinentes | |
| D,A | LAFLAMME C ET AL: "Harmonic-stochastic excitation (HSX) speech coding below 4 kbit/s" 1996 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING CONFERENCE PROCEEDINGS (CAT. NO.96CH35903), 1996 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING CONFERENCE PROCEEDINGS, ATLANTA, GA, USA, 7-10 M, pages 204-207 vol. 1, XP002106085 ISBN 0-7803-3192-3, 1996, New York, NY, USA, IEEE, USA * figure 1 * * alinéa 2.2 * ----- | 1 |
| | | DOMAINES TECHNIQUES RECHERCHES (Int.CL.6) |
| | | |
| Date d'achèvement de la recherche | | Examineur |
| 16 juin 1999 | | Krembel, L |
| CATEGORIE DES DOCUMENTS CITES | | |
| X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire | | |
| T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant | | |

2

EPO FORM 1503 03.92 (P04C12)

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)